

ALCOHOL OXIDASE 1 REGULATORY NUCLEOTIDE SEQUENCES  
FOR HETEROLOGOUS GENE EXPRESSION IN YEAST

Field of the Invention

The current invention is generally directed toward  
5 isolated polynucleotides comprising regulatory nucleotide  
sequences, vectors and expression cassettes containing such  
regulatory sequences, and host cells comprising the vectors  
and/or expression cassettes. In particular, the invention  
relates to 5' upstream regulatory nucleotide sequences  
10 within the promoter region of the alcohol oxidase 1 gene,  
vectors and expression cassettes containing such regulatory  
sequences, and host cells comprising the vectors and/or  
expression cassettes.

Background of the Invention

15 Methylotrophic yeast are capable of utilizing methanol  
as their sole energy source. The methylotrophic yeast  
comprise two groups, the asporogenous consisting of the  
*Candida* and *Torulopsis* species, and the ascomycetous  
consisting of the *Hansenula* and *Pichia* species. These yeast  
20 possess a conserved methanol utilization pathway. In the  
first step of this pathway, the enzyme alcohol oxidase  
catalyzes the oxidation of methanol to formaldehyde.  
Concomitantly, this reaction also generates hydrogen  
peroxide, a substance that is highly toxic to most cellular  
25 organelles. In order to avoid hydrogen peroxide toxicity,  
methanol metabolism is highly compartmentalized in  
specialized organelles, called peroxisomes, which sequester

this toxic byproduct from the rest of the cell. Alcohol oxidase has a low affinity for its substrate, oxygen. The partitioning of alcohol oxidase in the peroxisome, therefore, serves the additional role of concentrating the enzyme and substrate, thereby compensating, at least in part, for its low oxygen affinity.

The regulation of methanol metabolism in yeast generally occurs at the level of transcription, and includes control of the synthesis and activation of corresponding enzymes, as well as their degradation. Synthesis of methanol metabolizing enzymes is induced by methanol, formaldehyde and formate, and repressed by both glucose and ethanol. As stated above, a key enzyme in methanol oxidation is alcohol oxidase, which is also controlled both positively and negatively at the transcription level.

Several genes encoding alcohol oxidase from *Pichia pastoris* (AOX1 and AOX2), *Candida boidinii* S2 (AOD1), and methanol oxidase from *Hansenula polymorpha* (MOX1) are well characterized. Methanol induction causes the rapid de novo synthesis of alcohol oxidase, which is accompanied by a corresponding increase in alcohol oxidase mRNA. For example, in methanol grown *P. pastoris* cells, AOX rapidly accumulates up to 30% of the total soluble protein.

Yeasts have been extensively employed for the production of heterologous proteins and offer several advantages over prokaryotic expression systems. For example, some eukaryotic proteins that are produced in

prokaryotic cells are either unstable or lack biological activity. Accordingly, in these cases, yeasts offer advantages over their prokaryotic counterparts which include an intracellular environment that is more conducive for correct folding of eukaryotic proteins. Additionally yeasts, unlike prokaryotic hosts, have the ability to glycosylate proteins, which is important for both the stability and biological activity of the protein.

*Saccharomyces cerevisiae* was the first, and remains the most commonly employed eukaryotic expression system because its genome and physiology have been extensively characterized. It is not always the optimal expression system, however, for the large-scale production of heterologous proteins because of plasmid loss during scale-up, hyperglycosylation, and low protein yields. Recently, methylotrophic yeast expression systems have been developed and offer several advantages over their counterpart, *S. cerevisiae*. For example, methylotrophic yeast expression systems achieve very high cell densities in a simple defined medium, have strong inducible promoters, and the availability of methods, host strains, and expression vectors that facilitate genetic manipulations.

In particular, the methylotrophic yeast, *P. pastoris*, has been extensively utilized for the production of heterologous proteins at the industrial scale. *P. pastoris* is a particularly suitable expression system for the production of proteins at the industrial scale because of

the presence of a unique promoter, AOX1, used to drive gene expression. The AOX1 promoter, as stated above, is derived from the methanol regulated alcohol oxidase I gene and is one of the most efficient and tightly regulated promoters known. For example, as delineated above, in methanol grown *P. pastoris* cells, AOX rapidly accumulates up to 30% of the total soluble protein.

In addition to the AOX1 gene, *P. pastoris* also harbors a second functional AOX gene, AOX2. AOX2 encodes a protein that shares approximately 97% homology, and has the same specific activity as its AOX1 counterpart. However, the 5' upstream regulatory regions of AOX1 and AOX2 do not have significant regions of homology, despite the fact that both genes are repressed during growth on glucose, derepressed during carbon limitation, and induced by growth on methanol as the sole carbon source. In addition, AOX2 induction is responsible for only 15% of total alcohol oxidase activity in methanol-grown cell cultures. Due to the limited accumulation of AOX2 relative to AOX1, the AOX1 promoter is more advantageous to employ in an expression system for large-scale protein production.

Although high-level expression of heterologous proteins has been achieved employing the AOX1 promoter from *P. pastoris*, very little is known about the molecular mechanisms involved in methanol induction, either in *P. pastoris* or methylotrophic yeasts in general. The promoter region for AOX1 has been identified (SEQ ID NO: 1), however,

regulatory sequences within this region have not been extensively characterized. See Stroman et al., U.S. Pat. No. 4,855,231. In addition, the promoter region for the AOX2 gene has also been identified (SEQ ID NO: 2), and  
5 consists of three *cis*-acting regulatory elements that have been characterized (one positive and the other two negative). See Cregg., U.S. Pat. No. 5,032,516. The positive *cis*-acting element, the AOX2 upstream activator site, is required for a response to transcriptional  
10 induction by methanol. The two negative *cis*-acting elements are responsible for repression of the AOX2 promoter. However, as delineated above, the 5' upstream regulatory regions of AOX1 and AOX2 do not have significant regions of homology and thus, the extensive characterization of AOX2  
15 regulation provides little insight into the mechanism by which the AOX1 promoter is regulated.

In yet a further attempt to characterize regulatory regions of the AOX1 promoter, its sequence was compared to the 5' upstream regulatory regions of an alcohol oxidase  
20 gene, ZZA (SEQ ID NO: 3), isolated from an uncharacterized *P. pastoris* strain. See Kumagai et al., U.S. Pat. No. 5,641,661. This strain also contains two copies of the AOX genes. The deduced amino acid sequence of ZZA revealed that 14 of the first 16 amino acids are identical to that of the  
25 AOX1 and AOX2 genes. However, similar to the 5' regulatory regions of AOX1 and AOX2, the 5' upstream regulatory regions

of ZZA and AOX1 share only 66% homology. The two promoters, thus, may be regulated by completely distinct mechanisms.

Highly conserved nucleotide sequences, (TTGNNNGCTTCCAANNNTGGT) (SEQ ID NO: 4) and CCNCTTTTGTG (SEQ ID NO: 5) have been found in the 5' flanking regions of alcohol oxidase, methanol oxidase, and dihydroxyacetone synthase genes in *P. pastoris*, *H. polymorpha* and *C. boidini* S2. See Kumagai et al., U.S. Pat. No. 5,641,661. It was postulated that these conserved regions were involved in binding methanol-specific trans-acting factors. However, in a study characterizing the AOX2 promoter, these sequences were not involved in transcriptional regulation of the AOX2 gene since deletion of these regions did not impact regulation of the AOX2 promoter (Ohi, et al. (1994) Mol.Gen Genet 243:489-99). Thus, although these regions may be highly conserved across methylotrophic yeasts, their impact on transcriptional regulation of the AOX genes remains to be fully elucidated.

Accordingly, a need exists to identify promoters and other regulatory nucleotide sequences for the controlled expression and/or high level expression of heterologous proteins in yeast. In particular, a need exists to provide new AOX1 promoters and regulatory nucleotide sequences for the controlled expression and/or high level expression of heterologous proteins in yeast.

#### Summary of the Invention

Among the several aspects of the invention therefore is provided an isolated polynucleotide comprising a regulatory region containing a nucleotide sequence less than about 1000 nucleotides long selected from the group consisting of:

5 (a) SEQ ID NO: 16, a fragment of SEQ ID NO: 16, the complement of SEQ ID NO: 16, or a fragment of the complement of SEQ ID NO: 16;

(b) a polynucleotide that hybridizes to the polynucleotide of (a) under conditions of high stringency;  
10 and

(c) a polynucleotide with at least 80% sequence homology to the polynucleotide of (a).

Another aspect provides an isolated polynucleotide comprising a regulatory region containing a nucleotide  
15 sequence less than about 1000 nucleotides long selected from the group consisting of:

(a) SEQ ID NO: 17, a fragment of SEQ ID NO: 17, the complement of SEQ ID NO: 17, or a fragment of the complement of SEQ ID NO: 17;

20 (b) a polynucleotide that hybridizes to the polynucleotide of (a) under conditions of high stringency; and

(c) a polynucleotide with at least 80% sequence homology to the polynucleotide of (a).

25 In yet another aspect is provided an isolated polynucleotide comprising a regulatory region containing a

nucleotide sequence less than about 1000 nucleotides long selected from the group consisting of:

(a) SEQ ID NO: 18, a fragment of SEQ ID NO: 18, the complement of SEQ ID NO: 18, or a fragment of the complement  
5 of SEQ ID NO: 18;

(b) a polynucleotide that hybridizes to the polynucleotide of (a) under conditions of high stringency; and

(c) a polynucleotide with at least 80% sequence  
10 homology to the polynucleotide of (a).

Still another aspect of the invention is provided an isolated polynucleotide comprising a regulatory region containing a nucleotide sequence less than about 1000 nucleotides long selected from the group consisting of:

(a) SEQ ID NO: 19, a fragment of SEQ ID NO: 19, the complement of SEQ ID NO: 19, or a fragment of the complement  
15 of SEQ ID NO: 19;

(b) a polynucleotide that hybridizes to the polynucleotide of (a) under conditions of high stringency;  
20 and

(c) a polynucleotide with at least 80% sequence homology to the polynucleotide of (a).

In yet a further aspect of the invention is provided an isolated polynucleotide comprising a regulatory region  
25 containing a nucleotide sequence less than about 1000 nucleotides long selected from the group consisting of:



(a) SEQ ID NO: 20, a fragment of SEQ ID NO: 20, the complement of SEQ ID NO: 20, or a fragment of the complement of SEQ ID NO: 20;

(b) a polynucleotide that hybridizes to the  
5 polynucleotide of (a) under conditions of high stringency;  
and

(c) a polynucleotide with at least 80% sequence  
homology to the polynucleotide of (a).

A further aspect of the invention provides an isolated  
10 polynucleotide comprising a regulatory region containing a  
nucleotide sequence less than about 1000 nucleotides long  
selected from the group consisting of:

(a) SEQ ID NO: 21, a fragment of SEQ ID NO: 21, the  
complement of SEQ ID NO: 21, or a fragment of the complement  
15 of SEQ ID NO: 21;

(b) a polynucleotide that hybridizes to the  
polynucleotide of (a) under conditions of high stringency;  
and

(c) a polynucleotide with at least 80% sequence  
20 homology to the polynucleotide of (a).

Still another aspect of the invention provides a  
recombinant vector comprising the polynucleotide of SEQ ID  
NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID  
NO: 20, and SEQ ID NO: 21. Host cells comprising the  
25 recombinant vector are also provided.

A further aspect provides an expression cassette  
comprising the polynucleotide of SEQ ID NO: 16, SEQ ID NO:

17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21. Host cells comprising the expression cassette are also provided.

Brief Description of the Drawings

5        These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying figures where:

      Figure 1 depicts the construction of expression vector, pAL2 and mutagenesis vector pMMI101. PAL2 was constructed by inserting *Eco* RI/*Sma* I fragment of pSAOH5 into pIB1 promoterless vector. The vector pMMI101 was constructed by inserting *Eco* RI/*Bam* HI fragment of pSAOH5 into *Eco* RI/*Bam*.

      Figure 2 depicts deletion analysis of the AOX1 promoter. The plasmids were inserted into GS115 *his4* locus. Single copy integrands were identified by Southern assay using *HIS4* gene as a probe. The cells were grown in MM (induced) and MME (repressed) conditions.  $\beta$ -Gal activity of the AOX1-lacZ fusion was measured and reported as percent of activity of methanol-grown cells containing pAL2. The TATA box is indicated by the black square, and +1 ATG is indicated by the upside down, black triangle.

      Figure 3 depicts results of the gel shift assay for fragment A (SEQ ID NO: 16). Lane 1, no protein added; lane 2-5, six mg protein from MM grown cells; lane 6, six mg protein from MME grown cells; lane 3, thirty fold cold specific DNA (fragment A (SEQ ID NO: 16)); lane 4, thirty

fold cold fragment C (SEQ ID NO: 18); lane 5 thirty fold *L. monocytogenes* DNA.

Figure 4 depicts results of the gel shift assay for fragment C (SEQ ID NO: 18). Lane 1, no protein added; lanes 2-5, protein from MM grown cells; lanes 7-9, protein from MME grown cells; lane 3 and 7, thirty fold cold specific competitor (fragment C (SEQ ID NO: 18)); lanes 4 and 6, thirty fold cold unspecific competitor (fragment A (SEQ ID NO: 16)); lanes 5 and 9, thirty fold cold unspecific *L. monocytogenes* DNA.

Figure 5 depicts results of the gel shift assay for fragment F (SEQ ID NO: 21). Lane 1, no protein added; lane 2, six mg protein from MM grown cells; lane 3, twelve mg protein from MM grown cells; lane 4, six mg protein from MME grown cells; lane 5, twelve mg protein from MME grown cells.

Figure 6 depicts the consensus sequences of methanol regulated promoters of methylotrophic yeast. Occurrence of TTCAA (A) (Kumagai et al., 1993) and GATAG (B) (Ohi et al., 1994) core consensus sequence in AOX1 promoter. C, Alignment of the MOX UAS2 with AOX1 promoter (Godecke et al., 1994). Minus numbers are given relative to +1 ATG codon, +/- defines plus/minus strand. The fragment letters containing the sequences are also given in the right column.

#### Abbreviations and Definitions

To facilitate understanding of the invention, a number of terms and abbreviations as used herein are defined below:

As used herein "isolated polynucleotide" means a polynucleotide that is free of one or both of the nucleotide sequences which flank the polynucleotide in the naturally-occurring genome of the organism from which the polynucleotide is derived. The term includes, for example, a polynucleotide or fragment thereof that is incorporated in a vector or expression cassette; into an autonomously replicating plasmid or virus; into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule independent of other polynucleotides. It also includes a recombinant chimeric polynucleotide that is part of a hybrid polynucleotide, for example, one encoding a polypeptide sequence.

As used herein "polynucleotide" and "oligonucleotide" are used interchangeably and refer to a polymeric (2 or more monomers) form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Although nucleotides are usually joined by phosphodiester linkages, the term also includes polymeric nucleotides containing neutral amide backbone linkages composed of aminoethyl glycine units. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged

linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.), those containing pendant moieties, such as, for example, proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide. Polynucleotides include both sense and antisense strands.

As used herein, "sequence" means the linear order in which monomers occur in a polymer, for example, the order of amino acids in a polypeptide or the order of nucleotides in a polynucleotide.

As used herein, the terms "complementary" or "complementarity" refer to the pairing of bases, purines and pyrimidines, that associate through hydrogen bonding in double stranded nucleic acid. The following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil. The terms as used herein include complete and partial complementarity.

As used herein, the term "hybridization" refers to a process in which a strand of nucleic acid joins with a complementary strand through base pairing. The conditions employed in the hybridization of two non-identical, but very similar, complementary nucleic acids varies with the degree

of complementarity of the two strands and the length of the strands. Thus the term contemplates partial as well as complete hybridization. Such techniques and conditions are well known to practitioners in this field.

5       As used herein, "expression cassette" means a genetic module comprising a gene and the regulatory regions necessary for its expression, which may be incorporated into a vector.

      As used herein, "secretion sequence" or "signal  
10 peptide" or "signal sequence" means a sequence that directs newly synthesized secretory or membrane proteins to and through membranes of the endoplasmic reticulum, or from the cytoplasm to the periplasm across the inner membrane of bacteria, or from the matrix of mitochondria into the inner  
15 space, or from the stroma of chloroplasts into the thylakoid. Fusion of such a sequence to a gene that is to be expressed in a heterologous host ensures secretion of the recombinant protein from the host cell.

      As used herein, "operably linked" means any linkage,  
20 irrespective of orientation or distance, between a regulatory sequence and coding sequence, where the linkage permits the regulatory sequence to control expression of the coding sequence.

      As used herein, "heterologous coding sequence" means  
25 any coding sequence other than the one that naturally encodes the alcohol oxidase 1 protein, or any homolog of the alcohol oxidase 1 protein.

As used herein, "regulatory sequence" or "regulatory region" as used in reference to a specific gene refers to the coding or non-coding nucleotide sequences within that gene that are necessary or sufficient to provide for the regulated expression of the coding region of a gene. Thus, the term encompasses promoter sequences, regulatory protein binding sites, upstream activator sequences and the like. Specific nucleotides within a regulatory region may serve multiple functions. For example, a specific nucleotide may be part of a promoter and participate in the binding of a transcriptional activator protein.

As used herein, "coding region" refers to that portion of a gene which codes for a protein. The term "non-coding region" refers to that portion of a gene that is not a coding region.

#### Description of the Preferred Embodiments

Applicants have identified regulatory nucleotide sequences that are part of the non-coding region of the alcohol oxidase 1 ("AOX1") gene. The AOX1 gene encodes the enzyme alcohol oxidase 1, which catalyzes the oxidation of methanol to formaldehyde (as stated in more detail above). The AOX1 promoter, is an inducible promoter, and is primarily induced by methanol and carbon starvation, and repressed in response to glucose and ethanol. Although the AOX1 promoter has been identified (SEQ ID NO: 1), prior to applicants discovery, 5' regulatory nucleotide sequences

within the promoter region had not been characterized. The regulatory regions discovered by applicants, as set-forth herein, may be employed to increase the expression of genes of interest in a variety of cells.

5        In order to identify the regulatory regions of the present invention, applicants divided the 1052 base pair AOX1 promoter (SEQ ID NO: 1) into six fragments, A (SEQ ID NO: 16), B (SEQ ID NO: 17), C (SEQ ID NO: 18), D (SEQ ID NO: 19), E (SEQ ID NO: 20), and F (SEQ ID NO: 21). Through a  
10 series of systematic deletions comprising both sequential deletions and dropout mutations, in conjunction with gel shift assays, regions of the AOX1 promoter (SEQ ID NO: 1) element involved in regulation and DNA-protein interactions were identified (as set-forth in more detail in the example  
15 section below). Based upon this deletion analysis, as depicted in Fig. 2, fragments A (SEQ ID NO: 16), B (SEQ ID NO: 17), C (SEQ ID NO: 18), E (SEQ ID NO: 20), and F (SEQ ID NO: 21) contain positive regulatory sites, while fragment D (SEQ ID NO: 19) contains a negative regulatory site. In  
20 addition, based upon gel shift assays, as depicted in Fig. 3 and Fig 4, it is believed that fragments A (SEQ ID NO: 16), and C (SEQ ID NO: 18) contain sequences for different DNA-binding proteins. It should be noted that, specific nucleotides within a regulatory region may serve multiple  
25 functions. For example, a specific nucleotide may be part of a promoter and participate in the binding of a transcriptional activator protein. Thus, while applicants



believe that A (SEQ ID NO: 16), B (SEQ ID NO: 17), C (SEQ ID NO: 18), E (SEQ ID NO: 20), and F (SEQ ID NO: 21) contain positive regulatory sites, and fragment D (SEQ ID NO: 19) contains a negative regulatory site, in addition to the  
5 stated function, each of these regulatory regions may serve multiple regulatory functions.

The present invention, therefore, encompasses an isolated polynucleotide comprising a regulatory region containing SEQ ID NO: 16, a fragment of SEQ ID NO: 16, the  
10 complement of SEQ ID NO: 16, or a fragment of the complement of SEQ ID NO: 16. In one embodiment the polynucleotide (or the fragment, or the complement of either the polynucleotide or fragment) is less than about 1000 nucleotides long, in another embodiment the polynucleotide is between about 4  
15 nucleotides to about 750 nucleotides long. More preferably, however, the polynucleotide is between about 250 nucleotides to about 750 nucleotides long. The fragments may be employed as probes for SEQ ID NO: 16 or related sequences. Also included are isolated polynucleotides that hybridize to  
20 SEQ ID NO: 16, or the complement of SEQ ID NO: 16 under conditions of high stringency. In one embodiment these polynucleotides are less than about 1000 nucleotides long, in another embodiment these polynucleotides are between about 4 nucleotides to about 750 nucleotides long. Such  
25 polynucleotides may be used as probes for SEQ ID NO: 16 or related sequences.

Still another embodiment provides an isolated polynucleotide comprising a regulatory region containing SEQ ID NO: 17, a fragment of SEQ ID NO: 17, the complement of SEQ ID NO: 17, or a fragment of the complement of SEQ ID NO: 17. In one embodiment the polynucleotide (or the fragment, or the complement of either the polynucleotide or fragment) is less than about 1000 nucleotides long, in another embodiment the polynucleotide is between about 4 nucleotides to about 750 nucleotides long. More preferably, however, the polynucleotide is between about 250 nucleotides to about 750 nucleotides long. The fragments may be employed as probes for SEQ ID NO: 17 or related sequences. Also included are isolated polynucleotides that hybridize to SEQ ID NO: 17, or the complement of SEQ ID NO: 17 under conditions of high stringency. In one embodiment these polynucleotides are less than about 1000 nucleotides long, in another embodiment these polynucleotides are between about 4 nucleotides to about 750 nucleotides long. Such polynucleotides may be used as probes for SEQ ID NO: 17 or related sequences.

In yet another embodiment, the invention encompasses an isolated polynucleotide comprising a regulatory region containing SEQ ID NO: 18, a fragment of SEQ ID NO: 18, the complement of SEQ ID NO: 18, or a fragment of the complement of SEQ ID NO: 18. In one embodiment the polynucleotide (or the fragment, or the complement of either the polynucleotide or fragment) is less than about 1000 nucleotides long, in

another embodiment the polynucleotide is between about 4 nucleotides to about 750 nucleotides long. More preferably, however, the polynucleotide is between about 250 nucleotides to about 750 nucleotides long. The fragments may be employed as probes for SEQ ID NO: 18 or related sequences. Also included are isolated polynucleotides that hybridize to SEQ ID NO: 18, or the complement of SEQ ID NO: 18 under conditions of high stringency (as defined herein). In one embodiment these polynucleotides are less than about 1000 nucleotides long, in another embodiment these polynucleotides are between about 4 nucleotides to about 750 nucleotides. Such polynucleotides may be used as probes for SEQ ID NO: 18 or related sequences.

An additional embodiment provides an isolated polynucleotide comprising a regulatory region containing SEQ ID NO: 19, a fragment of SEQ ID NO: 19, the complement of SEQ ID NO: 19, or a fragment of the complement of SEQ ID NO: 19. In one embodiment the polynucleotide (or the fragment, or the complement of either the polynucleotide or fragment) is less than about 1000 nucleotides long, in another embodiment the polynucleotide is between about 4 nucleotides to about 750 nucleotides long. More preferably, however, the polynucleotide is between about 250 nucleotides to about 750 nucleotides long. The fragments may be employed as probes for SEQ ID NO: 19 or related sequences. Also included are isolated polynucleotides that hybridize to SEQ ID NO: 19, or the complement of SEQ ID NO: 19 under

conditions of high stringency. In one embodiment these polynucleotides are less than about 1000 nucleotides long, in another embodiment these polynucleotides are between about 4 nucleotides to about 750 nucleotides. Such  
5 polynucleotides may be used as probes for SEQ ID NO: 19 or related sequences.

In yet another embodiment, the invention encompasses an isolated polynucleotide comprising a regulatory region containing SEQ ID NO: 20, a fragment of SEQ ID NO: 20, the  
10 complement of SEQ ID NO: 20, or a fragment of the complement of SEQ ID NO: 20. In one embodiment the polynucleotide (or the fragment, or the complement of either the polynucleotide or fragment) is less than about 1000 nucleotides long, in another embodiment the polynucleotide is between about 4  
15 nucleotides to about 750 nucleotides long. More preferably, however, the polynucleotide is between about 250 nucleotides to about 750 nucleotides long. The fragments may be employed as probes for SEQ ID NO: 20 or related sequences. Also included are isolated polynucleotides that hybridize to  
20 SEQ ID NO: 20, or the complement of SEQ ID NO: 20 under conditions of high stringency. In one embodiment these polynucleotides are less than about 1000 nucleotides long, in another embodiment these polynucleotides are between about 4 nucleotides to about 750 nucleotides long. Such  
25 polynucleotides may be used as probes for SEQ ID NO: 20 or related sequences.

A further embodiment of the present invention provides an isolated polynucleotide comprising a regulatory region containing SEQ ID NO: 21, a fragment of SEQ ID NO: 21, the complement of SEQ ID NO: 21, or a fragment of the complement of SEQ ID NO: 21. In one embodiment the polynucleotide (or the fragment, or the complement of either the polynucleotide or fragment) is less than about 1000 nucleotides long, in another embodiment the polynucleotide is between about 4 nucleotides to about 750 nucleotides long. More preferably, however, the polynucleotide is between about 250 nucleotides to about 750 nucleotides long. The fragments may be employed as probes for SEQ ID NO: 21 or related sequences. Also included are isolated polynucleotides that hybridize to SEQ ID NO: 21, or the complement of SEQ ID NO: 21 under conditions of high stringency. In one embodiment these polynucleotides are less than about 1000 nucleotides long, in another embodiment these polynucleotides are between about 4 nucleotides to about 750 nucleotides long. Such polynucleotides may be used as probes for SEQ ID NO: 21 or related sequences.

As is well known in the art, stringency is related to the  $T_m$  of the hybrid formed. The  $T_m$  (melting temperature) of a nucleic acid hybrid is the temperature at which 50% of the bases are base-paired. For example, if one the partners in a hybrid is a short oligonucleotide of approximately 20 bases, 50% of the duplexes are typically strand separated at the  $T_m$ . In this case, the  $T_m$  reflects a time-independent

equilibrium that depends on the concentration of oligonucleotide. In contrast, if both strands are longer, the  $T_m$  corresponds to a situation in which the strands are held together in structure possibly containing alternating duplex and denatured regions. In this case, the  $T_m$  reflects an intramolecular equilibrium that is independent of time and polynucleotide concentration.

As is also well known in the art,  $T_m$  is dependent on the composition of the polynucleotide (e.g. length, type of duplex, base composition, and extent of precise base pairing) and the composition of the solvent (e.g. salt concentration and the presence of denaturants such as formamide). An equation for the calculation of  $T_m$  can be found in Sambrook et al. (*Molecular Cloning*, 2nd ed., Cold Spring Harbor Press, 1989) and is:

$$T_m = 81.5^{\circ}\text{C} - 16.6(\log_{10}[\text{Na}^+]) = 0.41(\% \text{ G} + \text{C}) - 0.63(\% \text{ formamide}) - 600/L$$
Where  $L$  is the length of the hybrid in base pairs, the concentration of  $\text{Na}^+$  is in the range of 0.01M to 0.4M and the G + C content is in the range of 30% to 75%. Equations for hybrids involving RNA can be found in the same reference. Alternative equations can be found in Davis et al., *Basic Methods in Molecular Biology*, 2nd ed., Appleton and Lange, 1994, Sec 6-8.

Methods for hybridization and washing are well known in the art and can be found in standard references in molecular biology such as those cited herein. In general, hybridizations are usually carried out in solutions of high

ionic strength (6X SSC or 6X SSPE) at a temperature 20-25°C below the  $T_m$ . High stringency wash conditions are often determined empirically in preliminary experiments, but usually involve a combination of salt and temperature that is approximately 12-20°C below the  $T_m$ . One example of high stringency conditions is 1X SSC at 60°C. Another example of high stringency wash conditions is 0.1X SSPE, 0.1% SDS at 42°C (Meinkoth and Wahl, *Anal. Biochem.*, 138:267-284, 1984). An example of even higher stringency wash conditions is 0.1X SSPE, 0.1% SDS at 50-65°C. In one preferred embodiment, high stringency washing is carried out under conditions of 1X SSC and 60°C.

In another embodiment, the present invention provides an isolated polynucleotide sequence with at least 80%, more preferably 90%, more preferably still 95%, even more preferably 97% and still more preferably 99% sequence homology to SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, or SEQ ID NO 21 or the complement of SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, or SEQ ID NO 21.

"Homology", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "homology" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Homology" can be readily

calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., *SIAM J Applied Math*, 48:1073 (1988). Methods to determine homology are designed to give the largest match between the sequences tested. Moreover, methods to determine homology are codified in publicly available programs. Computer programs which can be used to determine identity/homology between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology*, 12: 76-80 (1994); Birren, et al., *Genome Analysis*, 1: 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.*, 215:403-410



(1990)). The well known Smith Waterman algorithm can also be used to determine homology.

The isolated polynucleotide comprising the regulatory region of the present invention, without being bound by any particular limitation, may comprise any combination of SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21 or a fragment of any of these sequences. Additionally, the regulatory region may comprise any combination of the complement of SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21 or a fragment of the complement of any of these sequences. The regulatory region, may also comprise any combination of isolated polynucleotides that hybridize under conditions of high stringency to SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21 or that hybridize under conditions of high stringency to the complement of SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21. Furthermore, the regulatory region, may comprise any combination of isolated polynucleotides with at least 80% sequence homology to SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21 or the complement of SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21. Accordingly, for example, in one embodiment of the present invention, the regulatory region comprises SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, from 1 to about 3 copies of

SEQ ID NO: 20, and SEQ ID NO: 21. In yet another embodiment, the regulatory region comprises SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 20, and SEQ ID NO: 21.

The polynucleotides comprising regulatory regions of the present invention may be isolated by any method generally known in the art for isolating non-coding regions of nucleic acid sequence. One such method comprises employing hybridization of a probe to a genomic library to detect shared nucleotide sequence and is detailed in, for example, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, (1989) and Ausubel et al., *Short Protocols in Molecular Biology*, 3rd ed., John Wiley & Sons (1995).

Also included are vectors comprising the isolated polynucleotide regulatory sequences of the invention or the complement thereof, as set-forth in detail above. Any vector suitable for propagation in yeast may be employed such as, for example, pIB1 (Sears et al., (1998) *Yeast* 14: 783-90), pPICZ (commercially available through Invitrogen, Carlsbad, California) vector series, and pPIC9K (commercially available through Invitrogen, Carlsbad, California). The vector may also be either a cloning vector or an expression vector. A cloning vector is a self-replicating DNA molecule that serves to maintain a DNA segment into a host cell. The most common type of cloning vectors are bacterial plasmids. An expression vector is a cloning vector designed so that a sequence inserted at a

particular site will be transcribed into mRNA and in addition may be translated into a protein. Both cloning and expression vectors usually contain nucleotide sequences that allow the vectors to replicate in one or more suitable host cells. In cloning vectors, this sequence is generally one that enables the vector to replicate independently of the host cell chromosomes, and also includes either origins of replication or autonomously replicating sequences. Various bacterial and yeast origins of replication are well known to those skilled in the art and include, but are not limited to, the pBR322 plasmid origin, and the 2 $\mu$  plasmid origin. Ausubel et al., ed., *Short Protocols in Molecular Biology*, 3rd ed., Wiley & Sons, 1995.

The isolated polynucleotide comprising regulatory sequences of the present invention may be inserted into the vector by a variety of methods. In yet a further embodiment, the isolated polynucleotide comprising regulatory regions operably linked to a heterologous coding region may be inserted into the vector. In the most common method, the sequence is inserted into an appropriate restriction endonuclease site(s) using procedures commonly known to those skilled in the art and detailed in, for example, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, (1989) and Ausubel et al., *Short Protocols in Molecular Biology*, 3rd ed., John Wiley & Sons (1995).

Expression and cloning vectors can and usually do contain a selection gene or selection marker. Typically, this gene encodes a protein necessary for the survival or growth of the host cell transformed with the vector.

- 5 Examples of suitable selection markers include, *HIS4*, *ARG4*, *ble*, and *kan* (Cregg et al., (2000) Mol. Biotech. 16: 23-52). The selection marker can have its own promoter so that expression of the marker occurs independent of the polynucleotide construct. The marker promoter can be either  
10 a constitutive or an inducible promoter.

- In a further embodiment, the vector of the present invention also typically will include a sequence that is homologous with the host's sequence at a particular gene loci, for example a portion of the *His4* gene or *AOX* gene  
15 (both yeast genes), to enable the vector to insert into the host's chromosome by homologous recombination. Preferably, the homologous sequence employed in the invention will be at least approximately 200 nucleotides in length and have homology to yeast genomic DNA. Applicants have found that  
20 such integration is desirable, particularly when the host is *P. pastoris*, because vectors that are not integrated tend to be unstable.

- More particularly, in another embodiment is provided a recombinant polynucleotide comprising any of the above  
25 described regulatory regions of the invention operably linked to a heterologous coding region. Such recombinant polynucleotides are commonly employed in cloning or

expression vectors (as detailed above), although other uses are possible. The heterologous coding region may encode any protein of interest suitable for expression in the host cell such as enzymes, hormones, regulatory proteins, structural proteins, or antigens (for use in inducing an immune response).

The isolated polynucleotides of the present invention may be part of an expression cassette that minimally comprises, operably linked in 5' to 3' direction, a polynucleotide comprising a regulatory region of the present invention, a heterologous coding region and a transcriptional termination signal sequence functional in a host cell. The expression cassette may comprise any of the isolated polynucleotide regulatory regions of the invention, as previously discussed. Additionally, the heterologous coding region may encode any protein of interest suitable for expression in the host cell such as enzymes, hormones, regulatory proteins, structural proteins, or antigens (for use in inducing an immune response).

In a further embodiment, the expression cassette can also comprise an operably linked targeting sequence, transit or secretion peptide coding region capable of directing transport of the protein produced to the desired location. The expression cassette may also further comprise a nucleotide sequence encoding a selectable marker and/or a purification moiety. Various methods have been devised for the addition of such affinity purification moieties to

proteins. Representative examples can be found in U.S. Patent Nos. 4,703,004, 4,782,137, 4,845,341, 5,935,824, and 5,594,115. Any method known in the art for the addition of nucleotide sequences encoding purification moieties can be used for example those contained in Innis et al., *PCR Protocols*, Academic Press (1990) and Sambrook et al., *Molecular Cloning*, 2nd ed., Cold Spring Harbor Laboratory Press (1989). In yet another embodiment, the expression cassette can comprise a sequence that is homologous with the host's sequence at a particular gene loci to enable the cassette to insert into the host's chromosome by homologous recombination.

Encompassed within the present invention are host cells transformed with any of the constructs described herein comprising the polynucleotide regulatory sequences of the invention. The host cell is preferably a yeast cell. Yeast are preferred hosts because their usage provides a means to regulate gene expression when the regulatory sequences of the invention are employed. As stated above, the regulatory regions of the invention are repressible in response to repressing carbon sources (e.g. glucose, ethanol, fructose, galactose, sucrose and glycerol) and inducible in response to non-repressing carbon sources (e.g. sorbital, mannitol, threhalose, and alanine) when subjecting the transformed yeast to carbon starvation. Accordingly, gene expression can be strictly regulated based upon the medium selected to grow the cells.

Even more preferably, however, the host will be a methylotrophic yeast cell. Still more preferably, the methylotrophic yeast cell is selected from the group of genera consisting of *Hansenula*, *Candida*, *Torulopsis* and  
5 *Pichia*. Even more preferably, the yeast cell is from *P. pastoris*. The use of methylotrophic yeast provides an additional means to control gene expression when employing the regulatory regions of the invention. Methylotrophic yeast, as opposed to yeast in general, are capable of growth  
10 on methanol as their sole carbon and energy source. In addition, as delineated above, methanol is a strong inducer of the regulatory sequences of the invention. For example, in methanol grown *P. pastoris* cells, AOX rapidly accumulates up to 30% of the total soluble protein. Thus, the use of  
15 methylotrophic yeast as host cells in the present invention provides a mechanism to both reliably control gene expression and at the same time, achieve a high rate of expression. Introduction of the construct into the host cell can be accomplished by any method known in the art.

20 The present invention also includes methods for the production of proteins from the above described host cells transformed with any of the constructs characterized herein containing the polynucleotide regulatory regions of the present invention operably linked to a gene encoding the  
25 desired protein. Proteins can be expressed in any suitable host cells, but are preferably expressed in yeast host cells for the reasons previously identified. Host cells are

genetically transformed to produce the protein of interest by introduction of an expression vector containing the nucleic acid sequence of interest. The characteristics of suitable cloning vectors and the methods for their introduction into host cells have been previously discussed. Methods for such protein production are known to those skilled in the art. (Davis et al., *Basic Methods in Molecular Biology*, Elsevier Science Publishing (1986); Ausubel et al., *Short Protocols in Molecular Biology*, 2<sup>nd</sup> Ed., John Wiley & Sons (1992)).

Host cells are grown under appropriate conditions to a suitable cell density to optimize expression of the protein. If the protein accumulates in the host cell, the cells are harvested by, for example, centrifugation or filtration. The cells are then disrupted by physical or chemical means to release the protein into the cell extract from which the protein can be purified. If the host cells secrete the protein into the medium, the cells and medium are separated and the medium retained for purification of the protein.

Larger quantities of protein can be obtained from cells carrying amplified copies of the sequence of interest. In this method, the sequence is contained in a vector that carries a selectable marker and transfected into the host cell or the selectable marker is co-transfected into the host cell along with the sequence of interest. Lines of host cells are then selected in which the number of copies of the sequence have been amplified. A number of suitable



selectable markers will be readily apparent to those skilled in the art. For example, the *sh ble* gene is widely used as a maker for co-amplification. The *sh ble* gene product confers resistance to the drug Zeocin in both *E. coli* and yeast.

Proteins recovered can be purified by a variety of commonly used methods, including, but not limited to, ammonium sulfate precipitation, immuno precipitation, ethanol or acetone precipitation, acid extraction, ion exchange chromatography, size exclusion chromatography, affinity chromatography, high performance liquid chromatography, electrophoresis, and ultra filtration. If required, protein refolding systems can be used to complete the configuration of the protein.

The polynucleotide encoding the regulatory sequences of the invention, as described in detail above, may be employed to increase the expression of genes of interest in a variety of host cells. However, applicants discovery, in addition to increased gene expression, may also be employed for a number of other functions. For example, the regulatory sequences of the invention may be employed in a research setting to further characterize promoter function and to study peroxisome biogenesis.

The detailed description set-forth above is provided to aid those skilled in the art in practicing the present invention. Even so, this detailed description should not be construed to unduly limit the present invention as

modifications and variation in the embodiments discussed herein can be made by those of ordinary skill in the art without departing from the spirit or scope of the present inventive discovery.

5 All publications, patents, patent applications and other references cited in this application are herein incorporated by reference in their entirety as if each individual publication, patent, patent application or other reference were specifically and individually indicated to be  
10 incorporated by reference.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to  
15 be construed as merely illustrative, and not imitative of the remainder of the disclosure in any way whatsoever.

#### Example

#### Materials and Methods

Strains and plasmids. Plasmid pSAOH5 (Tschopp et al.,  
20 (1987) Nucleic Acids Res. 15: 3859-76) and *P. pastoris* GS115 (Tschopp et al., (1987) Nucleic Acids Res. 15: 3859-76) (*his4*) strain were the generous gift of J. M. Cregg, Keck Graduate Institute, Claremont, CA. pIB1 promoter less expression vector was provided by B. S. Glick (University of  
25 Chicago, Chicago, IL). *E. coli* strains JM109 (*endA1*, *recA1*, *gyrA96*, *thi*, *hsdR17*(*r<sub>k</sub><sup>-</sup>*, *m<sub>k</sub><sup>+</sup>*), *relA1*, *supE44*, *l<sup>-</sup>*, *D(lac<sup>-</sup>proAB)*, [*F'*, *traD36*, *proA<sup>+</sup>B<sup>+</sup>*, *lacI<sup>q</sup>ZDM15*] and ES1301 mutS (*LacZ53*,

mutS201::Tn5, thyA36, rha<sup>-</sup>5, metB1, deoC, IN (rrnD-rrnE), pALTER<sup>®</sup>-1 plasmid were purchased as a components of Altered Sites<sup>®</sup> II in vitro Mutagenesis System (Promega, Madison, WI).

5 Media, growth conditions and enzymatic analysis.  
Yeasts were grown in minimal glycerol media (MGY: 1.34% yeast nitrogen base without amino acids, 1% glycerol,  $4 \times 10^{-5}$  % biotin) and transformed cells were selected on Minimal Dextrose (MD) plates (1.34 % yeast nitrogen base without  
10 amino acids (YNB), 2% glucose,  $4 \times 10^{-5}$  biotin and 1.5% agar). If required 0.0004% histidine was added to the media.

The cells were grown on Minimal Methanol (MM, 1.34 % YNB,  $4 \times 10^{-5}$  biotin, 0.5 % methanol) and Minimal Methanol-Ethanol (MME, 1.34 % YNB,  $4 \times 10^{-5}$  biotin, 0.5% methanol and  
15 0.5% ethanol) until the optical density (600 nm) reached to 1.5-2.0 at 30° C.  $\beta$ -Gal activity was assayed using ONPG (o-Nitrophenyl B-D-glactopyranoside) as reported by Miller, J. H. Assay of  $\beta$ -galactosidase. in Miller, J.H. (ed), *Experiments in Molecular Genetic*, Cold Spring Harbor  
20 Laboratory, pp. 352-355 (1972) incorporating the modification proposed by Guarente and Ptashne, Fusion of *Escherichia coli* lacZ to the cytochrome c gene of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U S A* 78, 2199-203 (1981). The cells were permeabilized with  
25 chloroform and sodium dodecyl sulfate (SDS) and the results were reported in Miller Units.

General DNA techniques. Restriction enzymes and primers were from New England Biolabs Inc. (Beverly, MA), Site Directed Mutagenesis (SDM) was performed with the Altered Sites® II in vitro Mutagenesis System from Promega (Madison, WI) according to the manufacturer's recommendations. Primers were purchased from Sigma/Genosys (The Woodlands, TX). Protocols for the system were followed as described in the technical manual. All other DNA manipulations were as previously described (Sambrook, et al., (1989) Molecular Cloning: A Laboratory Manual). A 1.5 Kb *HIS4* fragment from pAL2 was used for Southern analysis to select single copy transformants.

Plasmid constructions. The plasmid pSAOH5 was first used as the parental plasmid to characterize the regulation of *LacZ* gene under the control of the *AOX1* promoter of *P. pastoris*. pAL2, an integrating plasmid, was constructed because of instability of the plasmid pSAOH5 in *P. pastoris*. The 3.6 kb *Eco* RI-*Nru* I fragment of pSAOH5 was inserted in the *Eco* RI-*Sma* I region of the pIB1, promoter less *Pichia* expression vector, and pAL2 parental plasmid was obtained. The *Eco* RI/ *Nru* I fragment of pSAOH5 contained the 5' untranslated region and first 15 amino acids of *AOX1* fused to the *lacZ* gene at codon nine. The mutagenesis plasmid, pMMI101, was constructed by inserting a 1120 base pair *Eco* RI-*Bam* HI fragment of pSAOH5 into pALTER-1 (Figure 1).

The plasmids containing 5' deletion derivatives of the AOX1 promoter were constructed using the Altered Sites® II in vitro Mutagenesis System to insert Eco RI sites at various positions along the promoter element. The primers used to obtain 5' deletion derivatives of the AOX1 promoter are as follows;

AOXRI132 5'-ACGCAGGAATTCCTCCACTC-3' (SEQ ID NO: 22)

AoxRi285 5'-GGCGAGGAATTCATGTTTGT-3' (SEQ ID NO: 23)

AoxRi424 5'-GTCTTGAATTCCTAATATGAC-3' (SEQ ID NO: 24)

10 AOXRI551 5'-GGTATTGAATTCACGAATGCTC-3' (SEQ ID NO: 25)

AOXRI699 5' -CTTCCAAGAATTCTGGTGGG-3' (SEQ ID NO: 26)

The mutant plasmids (pMMI132-pMMI669) were digested with Eco RI/Sma I, and the fragments were exchanged with the native AOX1 promoter in pAL2.

15 The construction of pALD series which remove internal AOX1 promoter sequences was as follows: pMMI series of plasmids, pMMI132, pMMI285, pMMI424 and pMMI551 which carry the mutant Eco RI sites and deletes the 5' upstream AOX1 promoter fragments, A (SEQ ID NO: 16), AB (SEQ ID NO: 16 and  
20 SEQ ID NO: 17, respectively), ABC (SEQ ID NO: 16, SEQ ID NO: 17, and SEQ ID NO: 18, respectively) ABCD (SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, and SEQ ID NO: 19, respectively), respectively. These fragments were inserted into the expression plasmids carrying the sequential  
25 deletion derivatives of the AOX1 promoter, pAL285, pAL424, pAL551, pAL669, and the plasmids were named pALD1, pALD2, pALD3 and pALD4, respectively. Consequently each pALD

plasmid lacks B (SEQ ID NO: 17), C (SEQ ID NO: 18), D (SEQ ID NO: 19) and E (SEQ ID NO: 20) promoter fragments respectively.

In order to purify the regions of the fragments (B (SEQ ID NO: 17), C (SEQ ID NO: 18), D (SEQ ID NO: 19) and E (SEQ ID NO: 20)) for band shift assay, a second round of site directed mutagenesis was done. The pMMI132, pMMI285, pMMI424, pMMI551 ssDNAs and AOXR1285, AoxRi424, AoxRI551 and AOXR1669 primers were used for SDM (site directed mutagenesis) reactions to drop B (SEQ ID NO: 17), C (SEQ ID NO: 18), D (SEQ ID NO: 19) and E (SEQ ID NO: 20) fragment, respectively. For example, to purify fragment B (SEQ ID NO: 17), pMMI132 single stranded DNA and primer AOX1285 were used to insert a third *EcoRI* site into the *AOX1* promoter. After sited directed mutagenesis, each of fragments A (SEQ ID NO: 16), B (SEQ ID NO: 17), C (SEQ ID NO: 18), D (SEQ ID NO: 19), E (SEQ ID NO: 20) and F (SEQ ID NO: 21) was gel purified with a Gene Clean II Kit (Bio101, Inc. Vista, CA).

Prior to transformation, the deletion and chimera c plasmids were digested with *Stu* I, which cuts the vectors once within the *HIS4* sequences to direct the integration event in the *his4* locus of the host GS115 strain.

Transformants were selected on MD media which lacks histidine. A single copy of the plasmids of transformed strains of transformed strains was determined by Southern blot assay. At least three independent single-copy transformants were chosen, induced in MM media, and  $\beta$ -Gal

activity was monitored in the strains. One representative strain was used for further studies. At least three independent induction experiments in MM and MME media were done to estimate the  $\beta$ -gal activity.

- 5 DNA probes for gel shift assays. Second round SDM (site directed mutagenesis) was done to purify plasmids pMMI132, pMMI285, pMMI424, pMMI551 to release fragments B (SEQ ID NO: 17), C (SEQ ID NO: 18), D (SEQ ID NO: 19) and E (SEQ ID NO: 20), respectively. The *Eco* RI fragment of pMMI132 and the
- 10 *Eco* RI/ *Sma* I fragment of pMMI669 were used as fragments A (SEQ ID NO: 16) and F (SEQ ID NO: 21), respectively. Each fragment was purified from agarose gel and end labeled with DNA polymerase using infrared labeled dATP (Li-Cor Inc, Lincoln, NE)
- 15 Preparation of yeast cell extracts. Yeasts were grown in 25 ml of MM or MME media at 30 °C to 1.5 OD (600 nm). The cells were centrifuged at 2,000Xg at room temperature, washed twice with 1M sorbitol and suspended in 500 ml of Buffer C [20 mM HEPES pH 8.0, 0.2 mM EDTA, 0.5 mM dithiotretiol
- 20 (DTT), 0.5 mM PMSF (phenylmethylsulfonyl fluoride), 0.42 M NaCl, 1.5 mM MgCl<sub>2</sub> and 25 % glycerol]. The cells were broken with glass beads (0.45-0.5 mm in diameter) in a bead beater (Biospec Products) (7X30 sec). After centrifugation at 19,000Xg for 15 minutes at 4 °C, the supernatant was used for
- 25 gel retardation assays. Total protein from cell extracts was estimated by the BCA assay (Pierce) using BSA as a standard protein.

Gel retardation assay. The binding reaction mixtures contained 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM DTT, 2 mg of poly (dI-dC)•poly (dI-dC) (deoxyinosine-deoxycytosine), 0.05% Nonidet-40, 10 % glycerol and end-labeled DNA in a volume of 20 ml. The binding reaction mixtures included 6 mg of protein extract. After 20 min of incubation at room temperature, mixtures were loaded on to a 6% non-denaturing pre-casted polyacrylamide gel (Novex, Carlsbad, CA). Electrophoresis was carried out at 10 v/cm of gel for 90 min in TBE (45 mM Tris/borate, 1 mM EDTA). The gels were dried and scanned with a prototype Y-axis scanner using Base ImagIR v. 4.0 software (Li-Cor, Lincoln, NE).

## Results

### Regulation of deletion derivatives of AOX1 promoter.

The plasmids carrying sequential deletions (pAL series) and dropout mutations (pDAL series) were integrated into the *his4* locus of chromosome of *P. pastoris* GS115 strain and single-copy transformants were screened by Southern blot analysis (results not shown). At least three single-copy transformants were screened by  $\beta$ -gal expression in MM medium, and one representative strain was used for further induction experiments. Each transformant was cultured in MG<sub>Y</sub> (minimal glycerol media) media overnight. The culture was used to inoculate MM and MME media, with cells were grown to an optical density of 1.5- 2.0 (600nm) prior to harvesting. The activity of the AOX1-lacZ fusion was measured and



reported as percent of activity of methanol-grown cells containing pAL2.

Each of the deletion derivatives of the AOX1 promoter decreased  $\beta$ -Gal activity in MM media compared to the native promoter (Fig 2). Loss of fragment D (SEQ ID NO: 19) (-518/-390) (pAL551) resulted in about a two-fold increase of  $\beta$ -gal activity, relative to pAL424, while loss of fragment E (SEQ ID NO: 20) (-390/-239) had the most severe effect on the expression.

10 The sequential deletions do not always reflect the relative promoter activity since the cumulative effect of cis-acting sequences is neglected. Therefore, a series of AOX1 promoter derivatives were constructed with internal deletions (pALDs) to determine the exclusive effect of each  
15 fragment. Only loss of sequences B (SEQ ID NO: 17) and E (SEQ ID NO: 20) had a significant effect on *lacZ* expression, suggesting that positive elements are located within these deletion intervals. The most significant effect on  $\beta$ -gal activity was observed by deletion of the fragment E (SEQ ID  
20 NO: 20) (-390/-239) where the AOX1 promoter lost 84% of activity.

Although the deletion derivatives of the AOX1 promoter had a significant effect on the promoter activity of methanol grown cells, the deletion and dropout plasmids did  
25 not impair ethanol induced repression of the AOX1 promoter (Fig 2).

Gel retardation assays. The GS115 (pAL2) strain was used to obtain total protein extracts growing in repressed (MME) or induced (MM) conditions. The fragments A, B, C, D, E and F (SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, and SEQ ID NO: 21, respectively) were tested in gel-shift assays. The fragment A (SEQ ID NO: 16) resulted in formation of specific DNA-protein complex when the cells were grown under induced condition (methanol) (Figure3, lane2). However, the ethanol grown cell-protein extracts did not reveal any DNA-protein complexes (Figure3, lane 6). Specificity was tested by adding unlabeled competitor and unlabeled nonspecific DNA (Lanes 3 and 5). Two mg poly dI·dC poly dI·dC and 500-bp *Listeria monocytogenes* gene product were used as nonspecific DNAs. The addition of competitor (unlabeled A fragment) resulted in loss of the signal (lane 3). Furthermore, addition of the unlabeled fragment C (SEQ ID NO: 18) (lane 4) did not compete with fragment A (SEQ ID NO: 16) implying that DNA binding for the fragment C (SEQ ID NO: 18) is specific and the protein binding to fragment A (SEQ ID NO: 16) does not form a complex with fragment C (SEQ ID NO: 18).

The gel shift assay using fragment C (SEQ ID NO: 18) and total protein extracted from yeast cells under induced and un-induced conditions revealed the formation of specific DNA-protein complexes. In this case the fragment A (SEQ ID NO: 16) and nonspecific DNA (*L. monocytogenes* DNA) did not

compete for the protein (Figure 4, lanes 4,8 and lanes 5, 9, respectively).

Extract from ethanol/methanol grown cells forms a complex with fragment C (SEQ ID NO: 18), but the intensity of the signal was weaker than that observed with methanol grown cells using the same amount of total cell extract (6 mg total protein).

The gel shift assays using the fragments B (SEQ ID NO: 17), D (SEQ ID NO: 19), E (SEQ ID NO: 20), and F (SEQ ID NO: 21) did not reveal any DNA-protein complexes, although different gel-shift experiment conditions were tested (results not shown). The gel shift results of the fragment F (SEQ ID NO: 21) is shown in Figure 5. Additionally, Figure 6 depicts consensus sequences of methanol regulated promoters of methylotrophic yeast.

In light of the detailed description of the invention and the examples presented above, it can be appreciated that the several aspects of the invention are achieved.

It is to be understood that the present invention has been described in detail by way of illustration and example in order to acquaint others skilled in the art with the invention, its principles, and its practical application. Particular formulations and processes of the present invention are not limited to the descriptions of the specific embodiments presented, but rather the descriptions and examples should be viewed in terms of the claims that follow and their equivalents. While some of the examples

and descriptions above include some conclusions about the way the invention may function, the inventor does not intend to be bound by those conclusions and functions, but puts them forth only as possible explanations.

5       It is to be further understood that the specific embodiments of the present invention as set forth are not intended as being exhaustive or limiting of the invention, and that many alternatives, modifications, and variations will be apparent to those of ordinary skill in the art in  
10 light of the foregoing examples and detailed description. Accordingly, this invention is intended to embrace all such alternatives, modifications, and variations that fall within the spirit and scope of the following claims.